Rapid detection and classification of aerosol events based on changes in particle size distribution

Phillip D. Stroud[*a], Christopher T. Cunningham[b], Gary Guethlein[b]
LAUR-00-1296, SPIE proc. vol. 4036, April, 2000.

[a]Los Alamos National Laboratory, MS-F607, Los Alamos, NM 87545
[b]Lawrence Livermore National Laboratory, 7000 East Ave, Livermore, CA 94550

ABSTRACT

A methodology is presented that allows aerosol particle size data to be used to indicate when an abnormal aerosol event may be occurring. Such data can be collected from an array of commercially available particle counter-sizers. The methodology employs two main elements: a detection element that recognizes when an aerosol concentration spike event is occurring; and a classification element that classifies aerosol events as normal (e.g. dust kicked up by wind gust or generated by normal vehicular activity) or abnormal (e.g. mistakenly released non-indigenous aerosol material). The detection element is based on observation of statistically significant rises in the aerosol concentration level, during an appropriate time interval. The classification element uses a new three dimensional feature space that highlights relevant differences in the aerosol particle size distribution function. The classifier adapts to the local environment by learning the region of the feature space that is occupied by normal aerosol events. Observations which then fall significantly outside this region are classified as abnormal. The methodology was developed using a set of atmospheric aerosol data containing over 600,000 observed aerosol particle size distributions, under both normal conditions, and with intentionally introduced abnormal aerosol. An implementation of the methodology is described. Many abnormal aerosol events in the data set are demonstrated to be distinguishable from normally occurring aerosol events.

Keywords: aerosol detection, feature space, adaptive algorithm, information extraction

I. INTRODUCTION

Every place has its own normally occurring aerosols. While there can be significant fluctuations in the aerosol constituents, concentration levels, and size distribution, the aerosol at a given location has some persistent characteristics. Much of the aerosol occurring at a location originates indigenously in the vicinity (e.g. particulate or pollen picked up by the wind or stirred up by local activity, or local industrial or vehicular aerosol sources). Other site-dependent components of the local aerosol depend on the prevailing atmospheric and meteorological conditions.

In a variety of situations, the presence of an abnormal aerosol -- an aerosol that does not belong to the locale—is indicative of danger or could be a cause for concern. In these situations, the capability to rapidly detect the presence of the abnormal aerosol could have immense value.

Among the simplest devices for characterizing aerosols in-situ is the light-scattering particle sizer-counter. These shoe-box sized, self-contained, commercially available (e.g. Met One Instruments, Inc.) sensors collect air through an intake port and pump it through a laser cavity. As aerosol particles pass through a small laser beam, a detector voltage, which depends on the detected intensity of scattered light, indicates the scattering cross-section, and thus the size, of the particle. In a matter of a few seconds, the sensor can determine the ambient concentration of aerosol particles in several size bins. This physical characterization is much faster than chemical or assay-based characterization.

[*] Correspondence: E-mail: stroud@lanl.gov; Telephone: 505-667-6654

In this paper, we present a methodology to detect abnormal aerosols using an array of particle counter-sizers. The raw concentration data is first filtered to remove measurement noise and irrelevant high-frequency concentration fluctuations. The data filtering process generates best estimates of time-averaged concentrations, with both short and long time-scales. A detection element is used to indicate the presence of an aerosol event. It responds to significant rises in the overall aerosol concentration that occur within appropriate time bounds.

Underlying the methodology is a transformation of the particle count data into a feature-space in which abnormal aerosols can be distinguished from normal aerosols. A classifier element indicates the likelihood that the aerosol is not a normal part of the local environment. The classifier responds to abnormalities in the aerosol particle size distribution.

Adaptation is an essential element of the methodology, for two reasons. First, the characteristics of normal aerosol vary greatly with place and time, so any detection method must adapt to the local background aerosol conditions. In addition, the characteristics of normal aerosol will undergo changes during operation, at time scales ranging from a half hour to weeks. The classifier has to adapt to these changes in order to maintain detection power without giving false positives.

The methodology is applicable to a single sensor, as well as to a detection system employing an array of sensors. For a multiple-sensor system, the abnormality of the aerosol is evaluated independently at each sensor, and a data fusion element is used to combine the sensor results. The data fuser looks for correlated events on different sensors, and includes a treatment of how the wind blows aerosols across the array.

In summary, the essential elements of the methodology are 1) data filtering to remove measurement noise and high frequency process noise, and to characterize recent and background concentrations 2) detection of concentration spike events, 3) transformation of filtered data to a feature space suitable for distinguishing relevant distribution function differences 4) characterization of the normal aerosol by adapting a partition in the feature space 5) determination of when a feature is sufficiently abnormal to trigger a detection, 6) data fusion of multiple sensors.

The methodology was developed using actual aerosol data. This data was collected simultaneously by twelve sensors, and consists of concentration measurements taken every 10 seconds in six different particle size bins. The data was collected over 25 nights, with typically seven hours of operation per night. Normal local conditions prevailed during much of the data collection, but there were also 75 instances when non-indigenous aerosols were intentionally introduced upwind of the sensor array. This data was instrumental in developing the feature space. It also allows the methodology to be assessed by the traditional performance measures of false positive rates and detection probabilities.

The particle size distribution function has been found to contain information that allows some abnormal aerosol events to be distinguished from normal aerosol events.

2. THE AEROSOL TEST DATA

Two data collection campaigns were conducted in the Utah desert, first during September of 1997, and again in September of 1998. Aerosol concentration measurements were made with an array of twelve particle counter-sizers (Met One Instruments, Inc.), distributed over a square mile of flat desert. Each sensor measured the concentration in six different size bins, at a rate of one measurement every 10 seconds. The data collection system was operated for a total of 185 hours, over 25 different nights. A total of approximately 2000 sensor-hours of data were collected, resulting in a database with more than 600,000 aerosol concentration profile records. In addition, the wind speed and direction was recorded every 10 seconds at each sensor location.

Normal local conditions prevailed during the bulk of the data collection, but there were also 75 instances when non-indigenous aerosols were intentionally introduced upwind of the sensor array. The normal local conditions include vehicular activity. This is the only existing data with multiple-particle-size aerosol concentration data from multiple sensors, that include both ambient conditions and intentional releases of non-indigenous aerosol.

The raw aerosol detector data consists of the number of particles counted during a sampling period, in each of six particle size bins. The factory pre-set particle size bins are 1 to 2µm, 2 to 4µm, 4 to 6µm, 6 to 8µm, 8 to 10µm, and 10µm and larger. For each measurement, the air is sampled for 9 seconds, at a volumetric flow rate of 0.1 cubic feet per minute. The particle counts are converted to concentrations (dividing counts by the 0.42475 liters per sample). The air inlets were located at a height of five feet above the ground.

As an example, the raw concentration data observed by sensor 6 is shown in Fig. 1, for the night of Sept. 15, 1998. This is just one out of 300 sensor-nights in the database, and will be used as a typical example in the rest of the paper. The time is labeled in hours of Sept. 15, with continuation after midnight, so that, for example, 2:00 a.m. on Sept. 16 is shown as 26:00 hours of Sept. 15. The data begins at 8:18 p.m. MDT on Sept. 15, 1998, and ends at 6:02 am, Sept. 16. There are a total of 3441 data records collected by sensor 6 on this night. The data from 60 measurements were lost during radio transmission from the sensor to the data collection point, but on no occasion were more than two consecutive data records dropped. The average aerosol concentration (of particles one micron and larger) was 466.0 particles per liter of air (pla), although instantaneous concentration values range from 40 to 40,000 pla. The average particle size distribution is as follows: 60.3% in the 1 to 2 µm, 35.4% in the 2 to 4 µm, 3.64% in the 4 to 6 µm, 0.417% in the 6 to 8 µm, 0.091% in the 8 to 10 µm, and 0.087% at 10 µm or larger. While the average over the whole night of the fraction of particles in the 1 to 2 µm size bin was 60.3%, for individual measurements, this fraction ranges from 43% up to 100%.

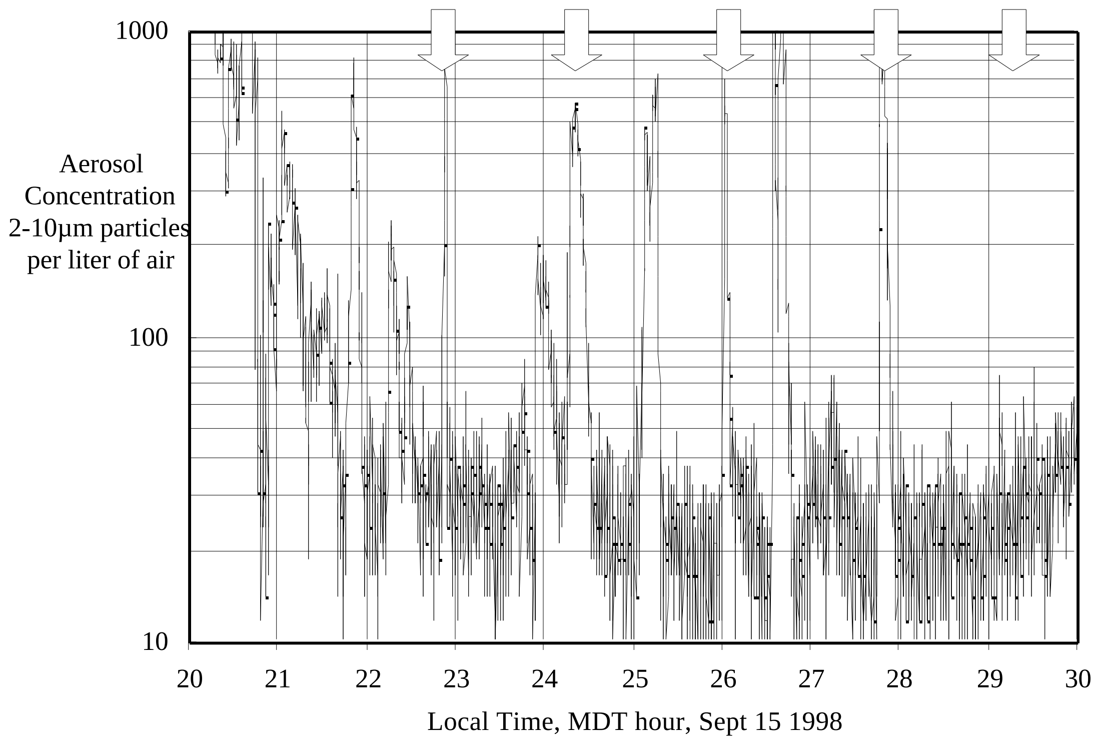


Figure 1. Raw data from Sensor 6, for the night of September 15-16, 1998, showing concentration of 2 to 10 µm aerosol particles. The arrows indicate the occurance of intentionally introduced non-indigenous ae rosol.

During the night of Sept. 15, 1998, there were five intentional releases of aerosol upwind from the sensor array. These releases are indicated by arrows in Fig.1. The aerosol concentration rise attributable to these releases is clearly visible in the sensor 6 data for the first four of these releases, but the fifth release barely registers. There are also many aerosol concentration spikes visible in Fig. 1, that were generated by vehicular activity and atmospheric phenomena. Similar behavior is seen in the data of all 12 sensors on all 25 nights. We look for a methodology that allows detection of as many intentional releases as possible, without giving false positives on the normally occurring spikes.

3. NOISE REMOVAL

The raw concentration data is corrupted by measurement noise and high-frequency process noise. Any discriminant will be more powerful if spurious noise can be removed from the data. The measurement noise is predominately shot noise, which is particularly significant for the larger size bins. The high-frequency process noise comes from small aerosol clouds, less that 30 meters in extent. These localized fluctuations in the aerosol will affect one observation, while not appearing in the previous or subsequent observation. It is assumed that these fluctuations do not carry information of interest.

Before the raw data is used, these two sources of noise are partially removed by filtering the data with recursive exponential filters. These filters have an adjustable time-scale, so that recent and background, or short, medium, and long time-scale filters can be implemented. An exponential filter with time-scale of τ gives a filtered value of $x(t) = \sum z(t')e^{-(t-t')/\tau} \Big/ \sum e^{-(t-t')/\tau}$, where z(t') is the raw concentration measurement taken at time t', and the sums extend over all data points collected prior to time t. This exponential filter has been implemented in an efficient recursive form. When a new data value is obtained, an update equation is used to revise the filtered value:

$$x_{n+1} = x_n + K_{n+1}(z_{n+1} - x_n)\,,$$

where the gain, K, is determined by the age of the previous update and the prior value of the gain:

$$K_{n+1} = K_n \Big/ \left(K_n + e^{-(t_{n+1} - t_n)/\tau}\right).$$

Prior to the first data point, the gain is set to unity, so that the filter will give the value of the first data point after it arrives.

A measure of validity of the filtered value can be derived from the gain. When the gain is high, the current filtered value is accorded a low confidence relative to the next observation. In that case, the current filtered value is not be considered a valid representative of the true underlying value. Each filter has an associated threshold value, K_{max}. Whenever the associated gain, K, exceeds this threshold, the filtered value is treated as invalid, and no decisions will be made in the algorithm using that value. The approach for setting K_{max} is based on the intuitive requirement that there must be N successive valid data before the filtered value will be considered. For data arriving every 10 seconds, the valid gain threshold is given by $K_{\max} = \left(1 - e^{-10t/\tau}\right) \Big/ \left(1 - e^{-10Nt/\tau}\right)$. This approach then accounts for data drop outs in a way that discounts older data appropriately. This recursive exponential filter has been found to give better results than simple moving averages or a non-adaptive Kalman filter.

4. DETECTION OF AEROSOL CONCENTRATION SPIKES

The task of detecting abnormal aerosol events can be broken into the two sub-tasks of 1) detecting aerosol events and 2) classifying them into normal and abnormal. A simple approach, using the 2 to 10 µm data is used to determine when an aerosol event is occurring at the sensor. Recursive exponential filters are used to define three filtered values: $x_{2\text{-}10}$, $xB_{2\text{-}10}$, and $\sigma^2_{2\text{-}10}$. The first of these, $x_{2\text{-}10}$, gives the recent concentration in the 2 to 10 µm particle size range. It is obtained with an exponential filter with a 60 second time scale, using the sum of the second through fifth size bin raw concentration data. This noise-removed concentration is shown in Fig. 2. Comparison with Fig. 1 shows that the aerosol spike events remain in the filtered data, while the high-frequency noise is greatly reduced. As with the raw data, concentration spikes occur during normal conditions, as well as when aerosols are intentionally introduced.

The second filtered quantity, $xB_{2\text{-}10}$, is a 4 minute exponential filter of 2 to 10 µm particles. The third filtered quantity, $\sigma^2_{2\text{-}10}$, keeps a running estimate of the variance of the difference between $x_{2\text{-}10}$ and $xB_{2\text{-}10}$. It is obtained with a fifteen minute recursive exponential filter of the square difference between recent and background concentrations:

$$\sigma^2_{210}(t) = \sum [x_{210}(t') - x^B_{210}(t')]^2 e^{-(t-t')/\tau} \Big/ \sum e^{-(t-t')/\tau}\,.$$

These three filtered values are used in a simple methodology to indicate when a concentration event is occurring. When the recent concentration exceeds the background concentration by a significant amount, a concentration event is occurring. The significance of the rise can be quantified by the parameter χ: $\chi = \left(x_{2-10} - x_{2-10}^{B}\right) / \sqrt{\sigma_{2-10}^{2} + 400}$. This parameter is plotted in Fig. 3, as a function of time, for the data collected by sensor 6 on September 15, 1998.

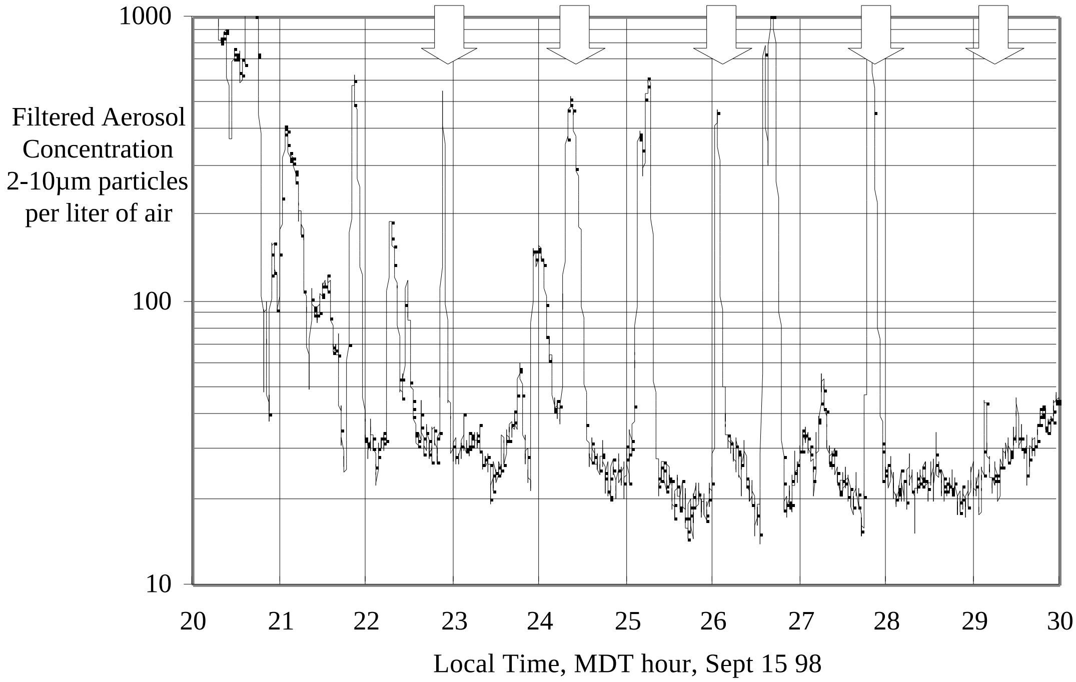


Figure 2. Recent concentration in the 2 to 10µm range, obtained with a 60 second exponential filtering of the raw data from sensor 6, on the night of September 15, 1998.

When this significance measure exceeds 1.5, an aerosol event is considered to be occurring. This condition can be specified as $x_{2-10} > x_{2-10}^{B} + 1.5\sqrt{\sigma_{2-10}^{2} + 400}$. This ensures that the recent concentration exceeds the longer-term concentration by a significant amount, with significance being measured in terms of the variance of this deviation. When the standard deviation (of recent concentration minus background concentration) is very small, this condition imposes a requirement that the recent concentration be at least 30 particles per liter above the background concentration. This prevents situations with small concentration rises following quiescent periods from generating spurious spike indications. The spike detector requires a significance of 1.5 times the standard deviation, in the more typical case wherein the standard deviation is much larger than 20 particles per liter of air.

There are two other conditions that must be met for the spike detector to indicate the presence of a spike. First, if these three filtered values are not all valid, the spike detector will not act. All three filtered values must be valid. The equivalent of eight consecutive valid data points are required for the recent and background concentrations to be valid, and the equivalent of ten consecutive valid recent and background concentrations must occur for the variance to be considered valid. The other condition is that $x_{2-10} > 100.0$ particles per liter of air.

For the ten nights of data collected in 1998, this spike detection formulation indicates the presence of a concentration spike after 8280 of the 302,704 total data records. The spike detector alone can thus reject over 97% of all observations as candidates for abnormal aerosol events. This is an important part of holding the false positive rate down.

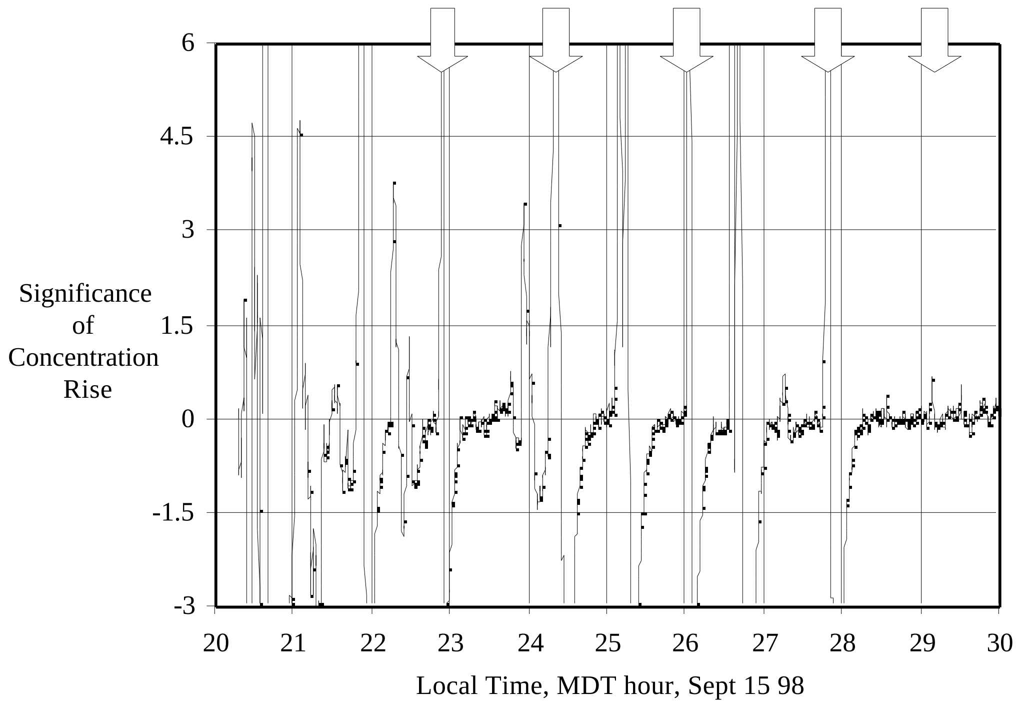


Figure 3. The significance of concentration rise, χ, for data from sensor 6, September 15, 1998.

A more sophisticated event detection methodology has also been developed, which makes two significant improvements on the simple formulation presented above. First, the concentration of the first principal component of the variation in the six particle size channels can be used instead of the concentration in the 2 to 10 µm range. Second, instead of fixing a hard threshold at 1.5 times the standard deviation, the statistics of the deviations can be accumulated and fit to a power law. The threshold can then be determined for a desired false positive rate. This more sophisticated method has been found to be superior to the simple method presented here, but its development was not completed.

5. REPRESENTATION OF PARTICLE SIZE DISTRIBUTION DEVIATIONS

Aerosol event detection alone is insufficient, since normal aerosol events produce spikes in the aerosol concentrations. A means of classifying normal aerosol events from abnormal aerosol events is thus required. Abnormal aerosols presumably could exhibit an arbitrary particle size distribution function, so the classifier can not simply look for certain distribution functions. A representation is required that highlights the difference of the particle size distribution from that of the local ambient background aerosol.

Although there are six particle size data channels, there appears to be only four meaningful channels. This can be seen by looking at the principal components of the variation of the six concentration channels. The principle components are the eigenvectors of the six by six covariance matrix. The first four principal components describe smooth trajectories in the 6-D concentration space, indicating the existence of a real underlying phenomena. The last two principle components, those with the smallest eigenvalues, are noisy, changing direction and eigenvalue rapidly. They therefore do not represent anything physically meaningful. We thus look for four information-bearing state variables, one that characterizes the overall concentration, and three more that characterize the shape of the particle size distribution function.

The noise is removed from the raw data with a four minute exponential filter, before the process of characterizing the size distribution function. Figure 4 shows these 4-minute filtered bin concentrations for the 2 to 4, 4 to 6, 6 to 10, and 10+ µm particle size bins. The 6 to 8 and 8 to 10 µm bins have been combined into one bin. The 1 to 2 µm data is ignored.

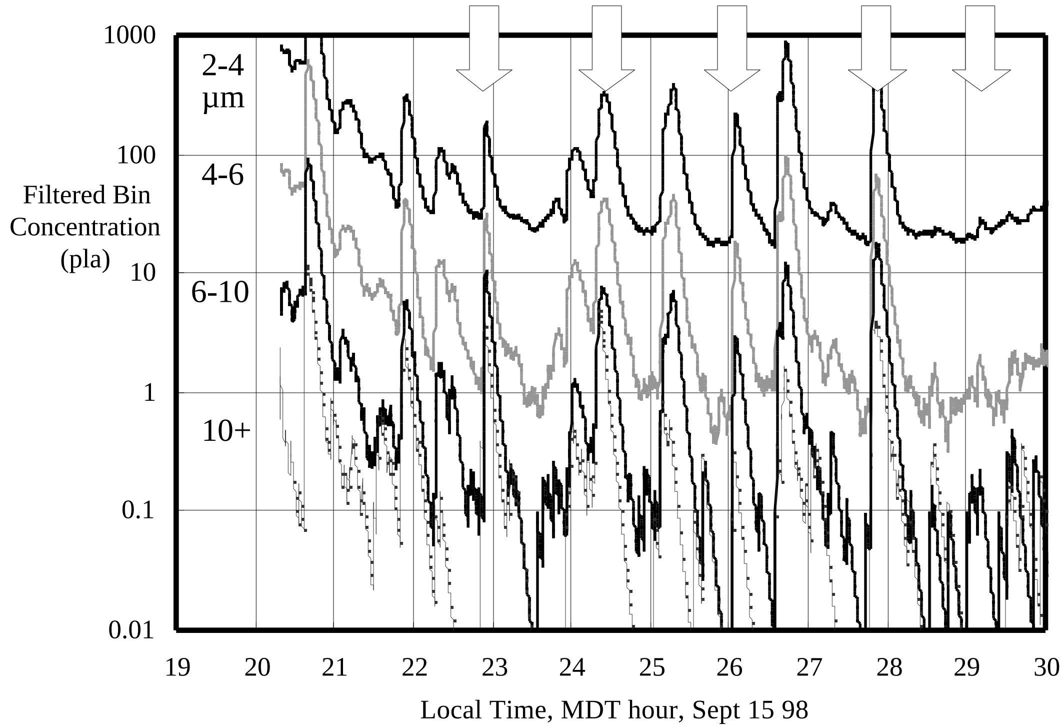


Figure 4. The 4-minute exponentially filtered aerosol concentration by particle size bin. The top line shows 2 to 4 µm. The second line down, shown as grey, gives the 4 to 6 µm range. The third line down, shown as heavy black, gives 6 to 10 µm. The lowest line shows 10 µm and larger. This data is from sensor 6, taken on the night of September 15, 1998.

When the observed concentration in one size bin is plotted against that of another size bin, a discriminant becomes apparent. Figure 5 shows the trajectory followed by the 4-minute filtered data in the 2-D plane defined by the concentration of 6 to 10 µm size particles, and the concentration of 4 to 6 µm size particles (again, illustrating with the sensor 6 data from September 15, 1998). An important observation is that the aerosols generated by vehicular activity tend to lie on the same lines as the ambient atmospheric aerosols, even though they show high concentration spikes. In this figure, two of the five intentionally released aerosols generate trajectories that are easily differentiated from those of normally occurring aerosol. Similar behavior can be observed in the 2-D plane formed by other pairs of size bins, and of course for other sensors and other nights.

Two simple bin-ratio formulations have been discovered that quantify differences in the bin-pair plane trajectories. A “slope” formulation characterizes, for example, the distribution function shape at around 6 µm by the ratio of concentration in the 6 to 10 µm range to the concentration in the 4 to 10 µm range: $s_6=x_{6\text{-}10}/x_{4\text{-}10}$. The “hat” formulation characterizes, for example, the distribution shape at around 8 µm by the ratio of concentration in the 6 to 10 µm range to the concentration in the 4 µm and larger range: $h_8=x_{6\text{-}10}/x_{4+}$. These feature variables take a value from 0 to 1. Fig 6 shows h_8 for the data collected by sensor 6 on September 15, 1998.

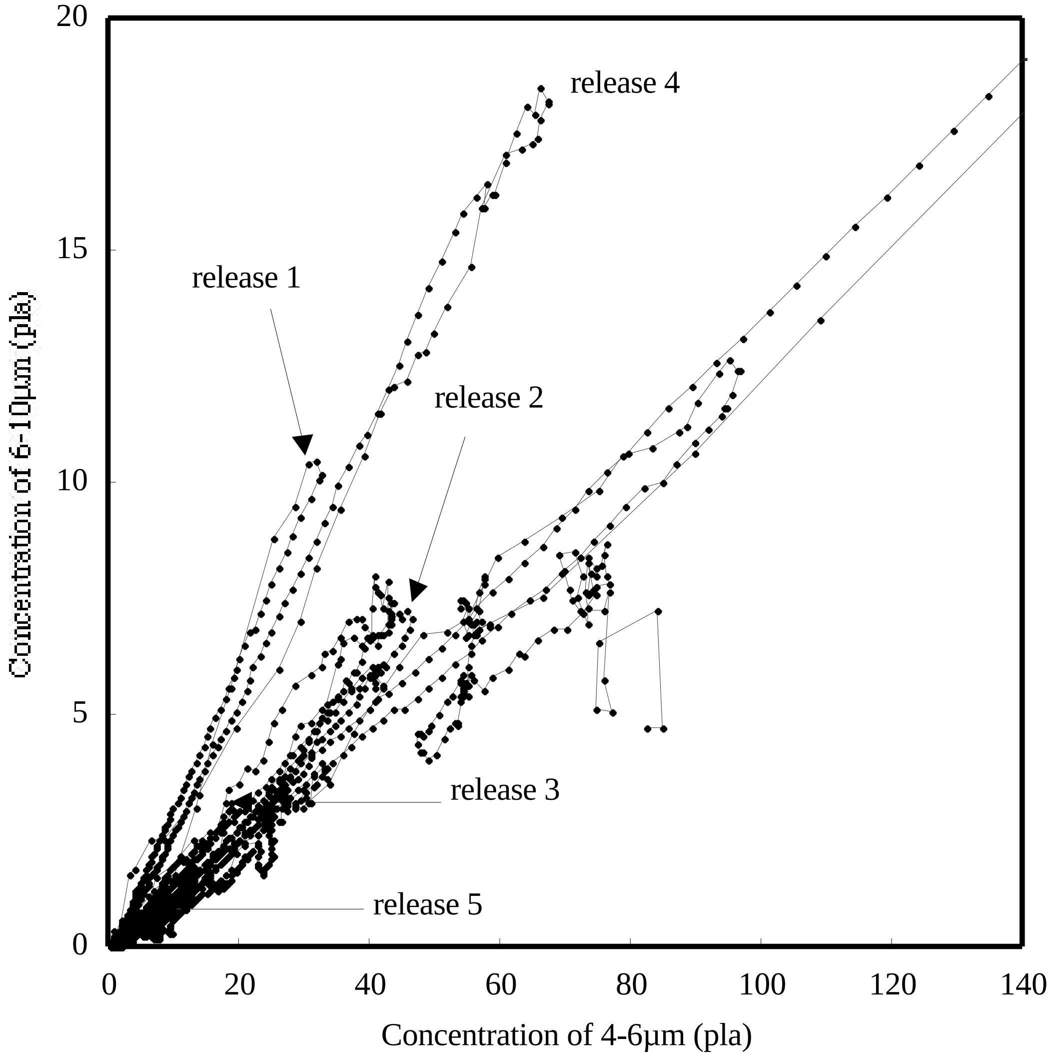


Figure 5. Trajectory described by the data from sensor 6, collected September 15, 1998, in the plane of 6 to 10 µm concentration vs. 4 to 6 µm concentration, using 4-minute exponential filter on the data

The first and fourth intentional aerosol releases stand out clearly in Fig. 6, because they attain values for h_8 that are significantly higher than any attained by normally occurring aerosols. There is one exception where the normal aerosol briefly shows a high value of h_8, but this spike is not accompanied by a rise in the total concentration level.

The shape of the distribution function can be characterized at other particle sizes, as well. We characterize the distribution at around 5 µm by using $h_5 = x_{4\text{-}6}/x_{2\text{-}10}$. Likewise, the shape of the distribution function at around 10 µm can be characterized by the feature variable $s_{10} = x_{6\text{-}10}/x_{6+}$. Figure 7 shows all data points in the feature space plane defined by h_8 and h_5, from sensor 6 on September 15, 1998. The points occupied by normal local aerosols are shown as gray circles, while the points generated by intentionally released aerosols are shown as black diamonds. It is apparent that some abnormal aerosols can be easily discriminated from normal aerosols based on differences in particle size distribution. In particular, the first and fourth releases on September 15, 1998 are clearly abnormal, as observed by sensor 6.

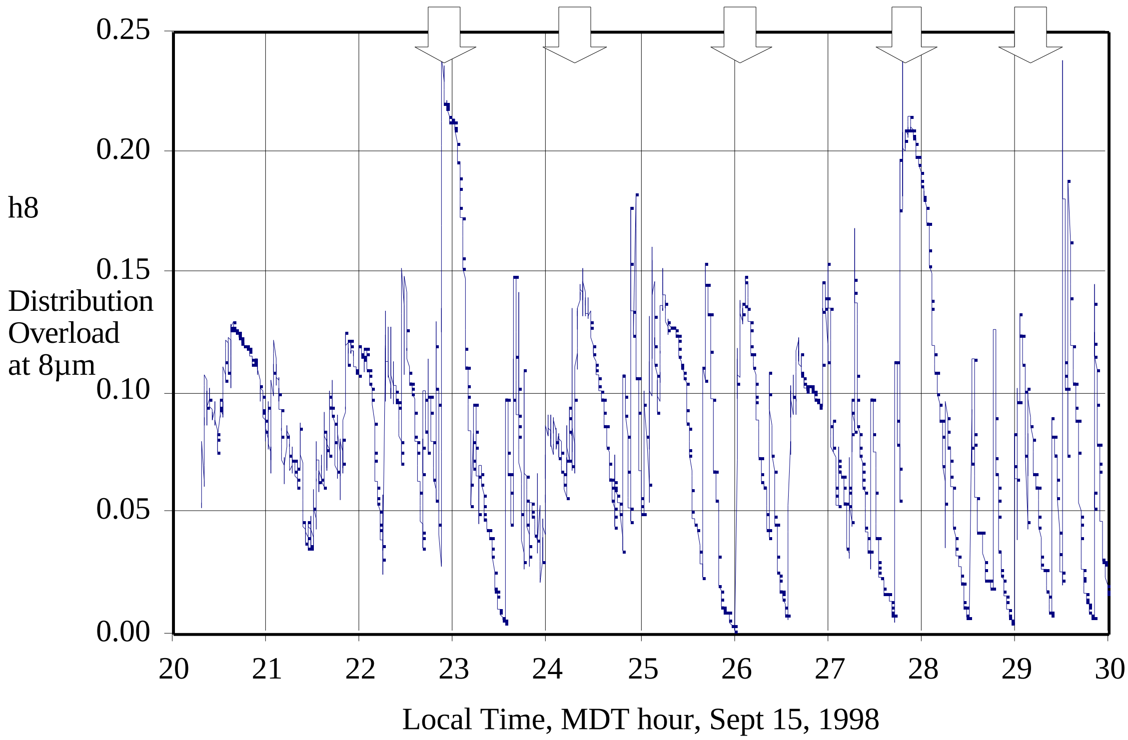
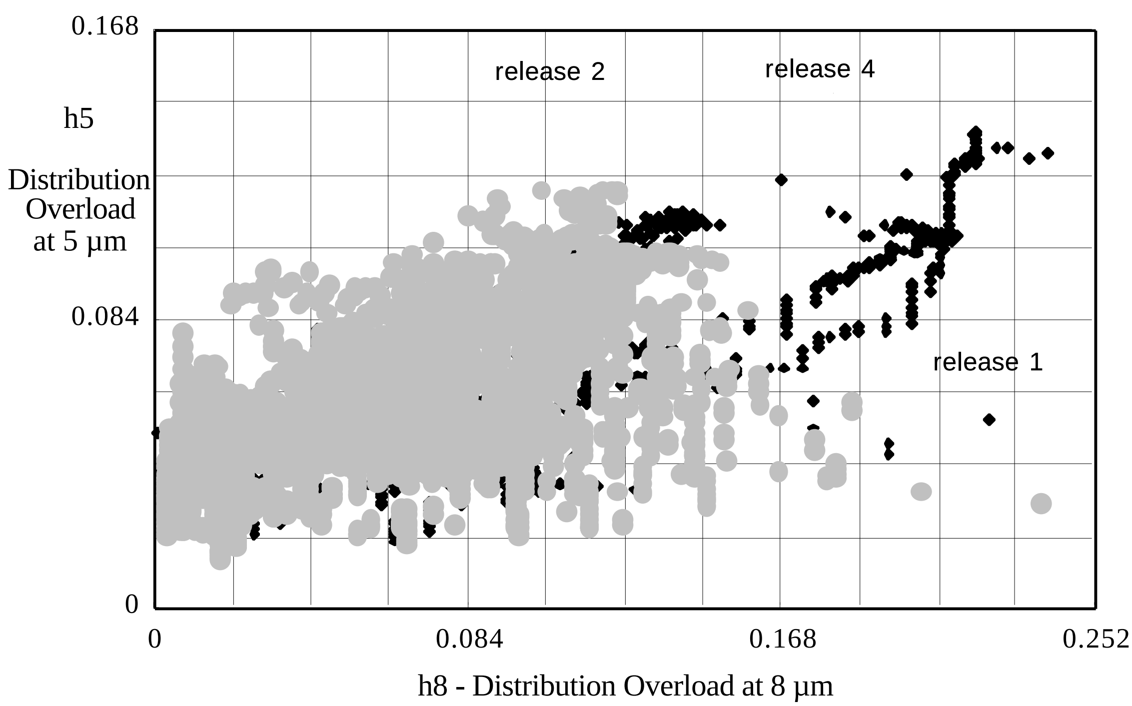


Figure 6. A derived feature variable, h_8, given by the ratio of 4-minute filtered concentration of particles in the 6 to 10 µm size range to that in the 4 µm and larger size range. Data shown is for sensor 6, Sept 15, 1998.



5. ALGORITHMS FOR DETECTING ABNORMAL AEROSOL

The details of implementation of detection algorithms based on this methodology are beyond the scope of this paper, but an overview of an example detection algorithm scheme is provided. The algorithm uses the spike detector described in section 3, and two classifiers, based on representations of the distribution function described in section 4. The slope-deviation classifier uses change in the slope characterization of the distribution at around 6 µm. The feature-space classifier identifies significant deviations in the 3-D feature space composed of the three feature variables h_5, h_8, and s_{10}.

For the feature-space classifier, each data record, after filtering, translates to a point in the 3-D feature space. There is a region in this 3-D feature space which is associated with normal aerosols. The boundary of this region forms a surface which can be described as a height in the s_{10} direction, as a function of location in the 2-D h_5-h_8 plane. Points beneath this boundary surface (i.e. lower values of s_{10}) are associated with normal aerosol.

The feature vector is evaluated only when the spike detector element indicates the presence of an aerosol event, and when the relevant filtered values are valid. When a feature vector occurs that is significantly outside of the normal aerosol region, it will be classified as an abnormal aerosol. The determination of whether a state vector is anomalous is based on a simple geometric distance in the 3-D state space, of the new vector from the boundary surface of the normal aerosol region. If this new point is below the normal aerosol region, or within a distance of 0.064 of it, the new point is considered too close to the historically observed values to be treated as anomalous. Otherwise, the new feature vector is declared to be anomalous, showing a completely different particle size distribution than has been seen by the sensor.

The normal aerosol region can not be updated immediately after data is received. This would be making use of information that will not be available in a deployed system: whether or not there is abnormal aerosol present. When a concentration observation is made, it may not be known for some time (e.g. 45 minutes), whether the aerosol is normal. Only if it is normal, should the data be used to modify the normal region of feature space. The data is used immediately to determine whether it is sufficiently anomalous to indicate abnormal aerosol. The feature vector derived from the data is then stored in a delay line. They pop out of the delay line after 45 minutes have elapsed. At that time, it can presumably be determined if the aerosol was normal, and whether to adapt the normal aerosol region in the feature space.

The boundary of the normal aerosol region in the 3-D feature space is continually adjusted as feature vectors pop out of the delay line. To a point (h_8, h_5, s_{10}) in the feature space, we define a block-shaped region consisting of that part of the feature space from the origin up to the point. All points with smaller values of all three variables are added to the normal aerosol region, if they are not already there.

When a sensor is initially set up, it will take some time to learn the region of the feature space that will be occupied by normal aerosols. Typically, a 10 hour training period is sufficient. This method of learning the normal aerosol region from data provides robustness against sensor calibration differences. When a sensor is shut down, the boundary of the normal aerosol region is saved to a file, so that it can be restored when the sensor is restarted.

The slope-deviation classifier looks for significant increases in the slope of the particle size distribution function at around 6 µm. The quantity $d_6 = x_{6\text{-}10}/x_{4\text{-}10} - (K\ xB_{6\text{-}10} - KB\ x_{6\text{-}10})/(K\ xB_{6\text{-}10} - KB\ x_{6\text{-}10})$ is used to characterize the change in the slope of the distribution function at around 6 µm, of the recent distribution relative to the background distribution. In this formulation, x gives the four-minute filtered concentrations, and xB gives a 30-minute background concentration. The gain of the four-minute filter, K, and the 30 minute filter, KB, are used to form a “lagged” background filter. The quantity d_6 is only evaluated if the recent and background filtered concentrations are valid, and the spike detector indicates the presence of an aerosol concentration spike. The denominators are obtained by adding the 4 to 6 micron concentration and the 6 to 10 micron concentration.

Several criteria must be satisfied before the slope-deviation classifier will indicate an abnormal aerosol. To ensure that the concentration values are statistically significant, the background concentration of 4 to 6 micron particles must be at least 2 particles per liter of air, and the recent concentration of 4 to 10 micron particles must be at least 15 particles per liter of air. Note that a concentration of 2 particles per liter of air would correspond to less than 1 aerosol particle count per sample. A second criteria ensures that the derivative at smaller particle size is consistent with that at 6 microns, while a third criteria causes the classifier to ignore aerosol events with exceedingly large fraction of very large particles, which are known to be local events, since the large particles would settle out from aerosols of distant origin.

When d_6 exceeds a threshold, the derivation deviation classifier will indicate that a non-ambient aerosol is present. The threshold is determined by adaptation to the local conditions. Each sensor keeps a threshold value d_{max}. Whenever a value d_6 is obtained that is larger than d_{max}, the value of d_{max} will be increased to d_6 if it is subsequently verified that the data were collected during normal aerosol conditions. As with the feature-space classifier, a 45 minute delay line is used for this purpose. The threshold is set at 0.02 when the sensors are first set up in a new environment. This approach gives a degree of automatic calibration. The actual threshold used to make the decision is a combination of the individual sensor threshold and the highest threshold currently in place over the whole array of sensors, designated $d_{highestMax}$. When d_6 exceeds $0.5 * (d_{max} + d_{highestMax})$ by a margin of 0.03, the aerosol is classified as abnormal.

The results of the feature-space classifier from each sensor are fused by the simple expedient of requiring wind-correlated detections of abnormal aerosol events from at least two sensors. The slope-deviation classifier will trigger on a single sensor. Both classifiers can effectively turn themselves off in environments where they would otherwise give false positives.

6. RESULTS

There were 75 intentional aerosol releases during the data collection. Of these, 11 occurred before the sensors had 10 hours of background data to train on. Of the remaining 64 releases, the algorithm described above was able to correctly classify 38 intentional release events as abnormal aerosols. In addition, the algorithm never misclassified any of the numerous normal aerosol event as abnormal.

7. LIMITATIONS

There are several circumstances in which this methodology would be ineffective for detecting the presence of abnormal aerosols. Since this approach requires a characterization of the normal aerosol, it wouldn't be effective for applications in which sufficient training time is unavailable. For the particular data set used to develop the methodology, it was found that a ten hour training period was required to sufficiently characterize the background region of the feature space. Other locales may require more or less time. If the normally occurring aerosol does not occupy a well-defined region of the feature space (as it did with the data employed here), the methodology would not work. If the abnormal aerosols of interest were to be intentionally tailored to match the normal aerosol, the methodology would not provide any classification power. When the abnormal aerosol is present in such small quantities as to disappear into the noise of the normal aerosol concentration, the spike detector would not be activated, and the methodology would never even look at the feature space discriminant. The detectable concentration of abnormal aerosol using implementations based on the methodology depends on the (presumably unknowable) distribution functions of the normal and abnormal aerosols, as well as the level of fluctuations in the normal aerosol concentration. Estimates of the expected sensitivity of the methodology can not therefore be reliably made without recourse to unjustifiable assumptions.

8. CONCLUSIONS

A new way of looking at the problem of abnormal aerosol detection has emerged, in which the task is divided into three simpler, understandable parts. The first part is to determine when an aerosol event is occurring. The second part is to determine if the particle size distribution is anomalous relative to what has been seen locally. The third is to fuse data from multiple sensors.

A very simple approach is used for aerosol event detection. It uses only the deviation of the recent concentration of 2 to 10 µm size particles from background levels, with significance based on the observed variance of this deviation. A more elaborate spike detection approach, using principle components of the variations, would perform somewhat better.

Transformation of the concentration by size bin into characterizations of the shape of the particle size distribution at several sizes is a novel and successful discovery of this effort. Several ratio formulations have been found to characterize the particle size distribution function at a given size. These ratios are very simple to implement, and are well suited to the multiple size bin form of the data. Both the hat formulation (i.e. the ratio of the concentration of 4 to 6 micron particles to the concentration of 2 to 10 micron particles) and the slope formulation (i.e. the ratio of the concentration of 4 to 6 micron particles to the concentration of 2 to 6 micron particles) are found to provide useful characterization of the particle size distribution around a particular particle size. Both of these ratio formulations have the advantage of ranging from 0 to 1. These ratios can be formed from recent bin concentrations and compared to the same ratios formed from background bin concentrations. Alternatively, the ratios formed from recent concentrations can be compared to the historical ratios seen previously at a sensor during normal aerosol conditions. These ratio formulations have a physical interpretation which provides a great advantage over more abstract mathematical constructs.

The adaptive capability of the algorithm is clearly essential to a system that can be taken anywhere at any time. The adaptive approach used by the algorithm appears to be successful. The fundamental idea of the adaptive approach is to discriminate abnormal from normal aerosols, while updating the region occupied by normal aerosols in a feature space. The observations are transformed into a feature space in which the discrimination is performed. The data corresponding to normal aerosols fall in a region of the feature space, which is gradually defined as data is collected. Detection of abnormal aerosol is based on the feature vector of an observation falling significantly outside this region. This adaptive approach allows classifiers to effectively shut themselves off in a location where they would give false positives. This provides a framework for extending the generality of the algorithm, by combining additional classifiers that focus on different features and discriminants. The effectiveness of classifying aerosol events according to whether the particle size distribution belongs to the locale, has been demonstrated for identifying abnormal aerosols.